IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA	)	
	)	
v.	)	Criminal No. 19-369
	)	
LAFON ELLIS		

Reply to Government's Response Concerning Cybergenetics' Violations of Protective Order and Request for Evidentiary Hearing to Resolve Material Issues of Fact.

I. Introduction.

The Court ordered the disclosure of various types of materials and stating that the source code must be produced "in a format allowing it to be reasonably reviewed, searched and tested." *See* ECF No. 161 at 12. The Court has recognized that this review and testing is necessary to determine whether the software used in this case is reliable when assessed according to the relevant disciplines, including computer science and software engineering. *See* ECF No. 151 at 1; *See also* ECF No. 138 (Denying government's motion to quash subpoena seeking access to the source code based, *inter alia*, on the Court's agreement that ' . . . independent source code review is critical when determining reliability. . . [and] Fundamental due process and fairness demand access,' citing *State v. Pickett*, 246 A.3d 279, 284 (N.J. Super. App. Div. 2021)).

Mr. Ellis and the government disagree on whether Cybergenetics violated the Court's order. It comes to no surprise that Cybergenetics is continuing to protract litigation by trying to argue their way out of having to produce materials necessary for reasonable inspection and testing. As a reminder to the Court, Cybergenetics vehemently objected to source code review and independent inspection of their software generated evidence over the course of the last year. Now that Cybergenetics has been forced to produce the source code, their new litigation tactic is to protract litigation through their refusal to provide materials covered by the protective order, including, but not limited to, producing the source code in a "format allowing it to be reasonably reviewed, searched and tested."

While Mr. Ellis understands that Cybergenetics' refusal to comply with this Court's order is not the government's fault, *per se*—as the subpoena is directed to a third party—the government is actively aiding the obstruction by continuing to serve as free counsel to a private company. In addition, instead of presenting arguments in the form of an expert affidavit contradicting Mr. Ellis' experts' claims that unproduced materials are covered within the meaning of the Court's order, the government, once again, engaged in *ad hominem*, attacks against Mr. Ellis' expert. *See* ECF No. 165 at 1.

While Mr. Ellis acknowledges that it could be argued that the relevant provisions at dispute between the parties in the Court's order are up for interpretation, it is Mr. Ellis' position that reasonableness of format of source code production conducive to testing should be interpreted by software engineering and computer science experts who are within the relevant scientific community of such review and testing. Interpreting some of these technology jargons through a software engineering lens is logical, reasonable, and within the spirit of this Court's decision that "independent source code review is critical when determining reliability. . . [and] Fundamental due process and fairness demand access." *See* ECF No. 138. Such a view is also consistent with the extensive briefing, amicus briefs, the Pickett decision, and expert declarations that this Court reviewed and considered in ordering inspection and testing of the code. *See* ECF No. 151 at 1.

To aid the Court's understanding of the issues presented in the attempted source code review, testing, and inspection, two of Mr. Ellis' experts, Dr. Jeanna Matthews, and Nathan Adams, wrote a declaration describing how the materials produced by Cybergenetics fail to comply with this Court's order and prevent reasonable inspection and testing. *See* Declaration of

Dr. Jeanna Matthews and Nathan Adams, attached as Exhibit 1. According to these software engineering experts,

> Cybergenetics has not provided key materials "in a format allowing it to be reasonably reviewed, searched and tested" as specified in the protective order issued on July 23 2021. . . . Of the two computers provided by Cybergenetics, the first, the one with an executable copy of TrueAllele and internet access, is still a black box. We cannot examine the inner workings of the executable system on that computer. The second computer is not a black box. Instead, it is a box filled with many small parts that are used in the building of TrueAllele. Imagine a box filled with an engine pulled apart into tiny parts – washers, bolts, pulleys, grommets, belts, valves and more – all mixed together, without schematics or instructions for how to construct the engine. Furthermore, we have reasons to believe that we *cannot* build a working version of TrueAllele because not all the necessary parts ("dependencies") are present (e.g. databases needed to run TrueAllele are missing). We intend to conduct a reasonable review, which is the ability to see inside the black box in order to inspect the contents of the working system. That access, via relevant software engineering materials including dependencies and build process documentation, has not been made available to us.

See Exhibit 1 at ¶ 2.

At this stage and given the protracted nature of the litigation so far, the defense asks the Court to intervene to clarify that Cybergenetics must disclose the materials that are necessary for review and testing. Considering the parties' disagreements on issues of material fact concerning obligations to produce materials necessary for reasonable review and testing, Mr. Ellis asks the Court to intervene by holding a hearing.

II. <u>Specific Replies to Government's Response at ECF No. 165.</u>

A. <u>Section 3(b) Violation</u>

In response to Mr. Ellis' claim that Cybergenetics is in violation of Section 3(b) of the protective order, which requires disclosure of, "[a]ll software dependencies including third-party code libraries, toolboxes, plugins, and frameworks," the government responded that the Court order "did not require Cybergenetics to produce databases and therefore they were not produced." *See* ECF No. 165 at 2. Mr. Ellis disagrees.

First, the term "all software dependencies," is software engineering and computer science technical language. Logically, by way of proper grammatical sentence structure, the sentence, "All software dependencies including third-party code libraries, toolboxes, plugins, and frameworks," includes more than just "third-party code libraries, toolboxes, plugins, and frameworks." The specific qualifier of "All software dependencies," includes every single software dependency. It also includes the specific enumerated items such as third-party code libraries, toolboxes, plugins, and frameworks.

A "dependency" is a module or component of a software system that provides functionality to the system, without which the system could not properly function. A database management system allows a user or software program to store and access data within a database. Database management systems (and the databases that they provide access to) are dependencies of software systems needing access to data stored in databases. PostgreSQL is a popular database management system which is used by TrueAllele to store and retrieve data.

The government concedes that databases are used by Cybergenetics when writing, "the executable program provided by Cybergenetics accesses the databases to operate." *See* ECF No. 165 at 2. That is, without access to the data stored in the database(s), TrueAllele could not properly function. Logically therefore, PostgreSQL itself and the database(s) accessed by TrueAllele through PostgreSQL, constitute dependencies of the TrueAllele software system. What follows is that without the database(s) installed and available on the *inspection* computer, it is reasonable to expect that TrueAllele's full functionality cannot be tested. As Dr. Matthews and Nathan Adams explained,

> Nor could the code be inspected via a debugger for system-level or many unit-level inspection and testing activities because dependencies such as databases were not included. Missing dependencies such as databases resulted in error states when attempting to run the TrueAllele source code provided, preventing system-level

> testing and many unit-level tests from being initiated and preventing execution path tracing from being conducted."

See Exhibit 1 at ¶ 7.

> we *cannot* build a working version of TrueAllele because not all the necessary parts ("dependencies") are present (e.g. databases needed to run TrueAllele are missing). We intend to conduct a reasonable review, which is the ability to see inside the black box in order to inspect the contents of the working system. That access, via relevant software engineering materials including dependencies and build process documentation, has not been made available to us.

See Exhibit 1 at ¶ 2.

B. Section 3(c) violation.

In response to Mr. Ellis' claim that Cybergenetics is in violation of Section 3(b) of the protective order, which requires disclosure of "Software engineering and development materials describing the development, deployment, and maintenance of the version(s) of the TrueAllele software system used in the instant case, including the software engineering documents recommended by organizations such as the Institute of Electrical and Electronics Engineers or the Internal Organization for Standardization," the government responded with a slew of arguments each of which Mr. Ellis will deconstruct.

The government first argued that, "[n]owhere in the Court order is there a requirement that Cybergenetics provide the defense expert instructions on how to re-create the program." *See* ECF No. 165 at 3. Mr. Ellis disagrees. As Dr. Matthews and Mr. Adams explain in their declaration:

> . . . build environment, materials, and instructions are relevant to our review and testing of software in general, including the TrueAllele software used in this case. Build instructions fall within "Software engineering and development materials describing the development, deployment of TrueAllele" and are used for inspecting and testing the source code. In order to confirm that the source code provided produces the TrueAllele software as it operated in this case, we must be able to build a functionally identical (if not entirely identical) version of TrueAllele from the source code provided. This requires build instructions.

See Exhibit 1 at ¶ 4. Dr. Matthews and Mr. Adam's statements concerning the necessity for "build instructions" for source code inspection and testing is supported by the Institute of Electrical and Electronics Engineers ("IEEE").

IEEE is the world's largest professional electrical, computer, systems, and software engineering organization with over 400,000 members in 160 countries.[1] IEEE produces "highly cited publications, conferences, technology standards, and professional and educational activities. IEEE is the trusted 'voice' for engineering, computing, and technology information around the globe."[2] Many of the IEEE standards have been adopted by U.S. governmental agencies including the Department of Defense[3], and the Nuclear Regulatory Commission[4].

This Court explicitly ordered that Cybergenetics produce IEEE recommended materials in section 3(c) of the protective order which states, "software engineering documents recommended by organizations such as the Institute of Electrical and Electronics Engineers [IEEE]" *See* ECF No. 161 at 3.

IEEE recognizes the need for "build instructions" for software testing. According to IEEE's Software Engineering Body of Knowledge (SWEBOK), version 3:

> Software building is the activity of combining the correct versions of software configuration items, using the appropriate configuration data, into an executable program for delivery to a customer or other recipient, such as the testing activity... Build instructions ensure that the proper build steps are taken in the correct sequence.

[1] Inst. of Elec. & Electronics Eng'rs, IEEE – About IEEE, https://www.ieee.org/about/index.html.

[2] *Id.*

[3] *See* U.S. Dep't of Def., Modeling & Simulation Coordination Office, *Modeling & Simulation Verification, Validation, & Accreditation Recommended Practices Guide Core Document* (2011).

[4] *See* U.S. Nuclear Reg. Comm'n, *Regulatory Guide* 1.168 – *Verification and Validation, Reviews, and Audits for Digital Computer Software Used in Safety Systems of Nuclear Power Plants* (1997).

See P. Bourque & R. E. Fairley, Eds., *Guide to the Software Engineering Body of Knowledge v3.0,* at 6.1, (2014), attached as Exhibit 2 at pg. 128.

Furthermore, IEEE's SWEBOK states:

> The build process and products are often subject to software quality verification. Outputs of the build process might be needed for future reference and may become quality assurance records.

See Exhibit 2 at Chapter 6.1 or pg. 128 of the Exhibit.

As evident from IEEE's SWEBOK, the "body of knowledge" within the software engineering profession recognizes the need for "build instructions" and the opines on that fact that build processes themselves are subject to software quality reviews under verification and validation activities. As such, materials describing those processes (i.e. instructions) are software engineering materials described under section 3(c) of the order.

Cybergenetics' refusal to provide "build instructions" is therefore not only violating section 3(c) of the protective order, but also section 9(c). Section 9(c) of the protective order states, "Any source code produced in discovery shall be made available for inspection, in a format allowing it to be reasonably reviewed, searched and tested. . ." As Dr. Matthews and Mr. Adams explained, software testing often involves building and executing software subsystems and subcomponents for incremental testing activities. Without build instructions describing how these components are appropriately configured and assembled, it can be practically impossible to ensure that "proper build steps are taken in the correct sequence," as required by IEEE's SWEBOK, to assemble, review, and test the version of TrueAllele used in this case.

The government next argues that "build instructions" should not be provided because in their view "[s]uch an order would give the defense expert complete and unfettered access to the program, which is not consistent with the verbiage of the Court order or the spirit or Order, which provides Cybergenetics some protection for its trade secrets. Under the order, as interpreted by the defense, there would be no protection." *See* ECF No. 165 at 3. Mr. Ellis disagrees by the governments baseless and unsupported assertion.

First, as argued above, "build instructions" are essential and were ordered produced by the Court. They did not have be specifically named because they fall within the category of materials to be produced under sections 3(c) and 9(e) of the order. Second, the government is correct, Mr. Ellis's experts need complete and unfettered access to the program to inspect and test it. Less than complete and unfettered access to the program can only serve to inhibit review and testing activities. Third, other than conjecture and baseless assertions, the government provides no factual basis, expert testimony, relevant publications or standards within the software engineering field to support their claim. Lastly, because they have no actual basis for why "build instructions" should not be produced, the government resorts to fearmongering the Court about Cybergenetics' trade-secret interests. Those interests were resolved through the issuance of a protective order by this Court, which representatives for Cybergenetics signed.

In another argument, the government tried to argue that "build instructions" are not needed because "Cybergenetics produced the source code, which demonstrates how the program was constructed. The source code is essentially the blueprints – the line-by-line instructions for exactly how this program was built and how it works." *See* ECF No. 165 at 3. Mr. Ellis disagrees. Thousands of MATLAB source code files were provided on the source code inspection PC. These constitute the instructions for TrueAllele's interconnected subcomponents but **do not** include explicit configurations or instructions for how TrueAllele was built from its components. As previously explained in more details above, and within the attached declaration at Exhibit 1, build instructions describe the process of integrating these separate components (and their dependencies) into a fully assembled system which is required for software testing and inspection.

In another argument on "build instructions" the government argues that "even if step-by-step instructions were available, this defense expert is not qualified to re-create a program of this degree of sophistication, and, even if he was, it would take years to do correctly." *See* ECF No. 165 at 3. This argument mischaracterizes what the experts in this case will do. Mr. Ellis' goal is not to reinvent TrueAllele from scratch. It is to have his experts review, test, and evaluate the TrueAllele software system used in this case by following software engineering principles and industry standards for said inspection and testing.

The government's final argument against producing "build instructions" is fundamentally nonresponsive. By stating "[i]t is therefore not clear how rebuilding a program advances a determination of its reliability when the defense expert can just test the version of the program used in this case as opposed to a version of the program that he would attempt to rebuild, undoubtedly unsuccessfully." *See* ECF No. 165 at 3. In their argument, the government asks the Court to substitute its lay opinion for Mr. Ellis' expert's, whom it has conceded is in fact an expert and yet ignores. What the government says is incorrect. As previously explained, the executable form of TrueAllele does not allow direct inspection of the software's internally executing instructions as it performs steps for mixture deconvolution or likelihood ratio calculation. Evaluation and testing of individual internal components of TrueAllele, such as for common software engineering activities of unit testing or tracing execution paths, requires access to the source code in a buildable and operable form. That is the entire point of this source code access litigation as explained in extensive briefing which the government seems to gloss over.

Lastly, the government's response motion at ECF No. 165 was nonresponsive to Mr. Ellis' claim that Cybergenetics violated section 3(c) of the protective order by failing to disclose "software engineering documents recommended by organizations such as the Institute of Electrical and Electronics Engineers or the Internal Organization for Standardization." The government provided zero argument as to why Cybergenetics is in compliance with these required disclosures, it can't, because Cybergenetics refused to disclose any such items.

C. <u>Section 3(d) violations.</u>

In response to Mr. Ellis' claim that Cybergenetics is in violation of Section 3(d) of the protective order, which requires Cybergenetics to produce "[a]ll records of unexpected results, including false inclusions, false exclusions and the conditions under which the unexpected results were achieved," *see* ECF No. 161 at 3, the government responded with that's "incorrect" followed by "to the extent that these records exist, they were produced." *See* ECF No. 165 at 4. Mr. Ellis claims that nothing was produced, zero. That's a problem. Is the government conceding that Cybergenetics is not in possession of any such records? The government's response is vague and nonresponsive and the Court should order them to clarify what Cybergenetics is, or is not, in possession of.

D. <u>Section 9(c) violation.</u>

Cybergenetics violated section 9(c) of the protective order which states "[a]ny source code produced in discovery shall be made available for inspection, in a format allowing it to be reasonably reviewed, searched and tested." *See* ECF No. 161 at 12. Cybergenetics is in violation of this provision because they did not produce the source code in a format allowing it to be "reasonably review[able], . . . and tested." *Id.* The government's response to this assertion was nonresponsive. Instead of countering Mr. Ellis' claims that Cybergenetics' refusal to provide "build instructions" and "databases" does not violate this Court's order that "[a]ny source code produced in discovery shall be made available for inspection, in a format allowing it to be reasonably reviewed, searched and tested," the government proceeded to give their own layman's opinion about what is, and is not, needed for source code review. *See* ECF No. 165 at 4 ("the defense has access to the actual software used in this case, and it seems to make a lot more sense to test that version of the software, as opposed to a version that the defense expert attempted to re-create."). Frankly, the government's own opinion that section 9(c) was not violated is irrelevant. Facts matter, and the fact is that according to the declaration of Dr. Matthews and Mr. Adams, two experts in the relevant field,

> Cybergenetics has not provided key materials "in a format allowing it to be reasonably reviewed, searched and tested" as specified in the protective order issued on July 23 2021.

See Exhibit 1 at ¶ 2.

> The form and materials on the source code inspection PC examined by Nathaniel Adams on August 18-20, 2021 do not allow for routine software inspection and testing tasks to be conducted on the TrueAllele source code provided.

See Exhibit 1 at ¶ 9.

The materials such as the "build instructions" and "databases" that Mr. Ellis' experts say are necessary for "routine software inspection" are normally produced in other cases where source code of probabilistic genotyping systems has been conducted. These are not materials outside the normal bounds of source code review. The developers of Cybergenetics main competitor STRmix provided their source code with these materials because that is the *standard way* of producing source code. As explained in the declaration to the Court,

> Past inspections of the probabilistic genotyping software programs Forensic Statistical Tool and STRmix by Nathaniel Adams involved the provision of build configurations and dependencies sufficient to establish functional equivalence between the source code provided for inspection and the versions of the programs

> used in their respective cases. The source code was operable within an IDE, enabling dynamic testing and execution tracing (via a debugger) for review purposes.

See Exhibit 1 at ¶ 8.

Cybergenetics is in violation of section 9(e) because it refuses to provide the source code "available for inspection, in a format allowing it to be reasonably reviewed, searched and tested," as required by the Court's order because it failed to provide the source code in a format deemed necessary by the IEEE, software engineering experts such as Dr. Matthews and Mr. Adams, and even in the manner that its biggest market competitor STRmix produced theirs.

III. <u>Conclusion and Requested Relief.</u>

Cybergenetics has violated multiple provisions of this Court's order. From the onset of this case, Cybergenetics obstructed access to independent verification and validation of their software generated evidence. After this Court ordered the source code disclosed, Cybergenetics again attempted to impede and delay the process by filing a motion to reconsider. When the Court denied Cybergenetics' motion to reconsider, Cybergenetics resorted to attempting to impede and obstruct reasonable access by means of prolonged litigation concerning the protective order. The litigation concerning the protective orders did not focus on Cybergenetics' attempts to protect their trade-secrets, instead they focused on Cybergenetics' objections to manner and method of source code production. After the parties finally reached an agreement to protective order language, Cybergenetics is again obstructing and impeding access by refusing to provide the source code in a format conducive to reasonable testing and inspection.

Cybergenetics may be afraid about their trade-secrets being disseminated—a fear that a protective order addresses—but what they are genuinely scared of is what independent review and testing of their source code will uncover about their software that's never been scrutinized.

The parties disagree about whether Cybergenetics violated the protective order. Mr. Ellis has clearly explained what is missing, how it was ordered produced, and supported his assertions through expert testimony and secondary sources relevant to the software engineering field. The government claims Cybergenetics are not in violation. Multiple factual disputes have arisen between the parties concerning issues of material fact on whether Cybergenetics is in compliance or violated the protective order. Based on these disputes, and because judicial intervention is needed, Mr. Ellis requests this Court to hold a hearing at which it can issue a ruling after hearing from all the relevant parties and their experts. In the words of John Henry Wigmore, cross-examination is "beyond any doubt the greatest legal engine ever invented for the discovery of truth." 3 Wigmore, Evidence §1367, p. 27 (2d ed. 1923).

Mr. Ellis requests a hearing, or an Order from the Court requiring Cybergenetics to produce the materials or face either exclusion of evidence or contempt or both.

Respectfully submitted,
/s/ Khasha Attaran
Khasha Attaran
Assistant Federal Public Defender